

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/865,018

DATE: 11/29/2001

TIME: 15:58:29

Input Set : A:\09865018.raw

Output Set: N:\CRF3\11292001\I865018.raw

SEQUENCE LISTING

ENTERED

```

3 (1) GENERAL INFORMATION:
5     (i) APPLICANT: Massague, Joan
6                   Roberts, James M.
7                   Koff, Andrew
8                   Polyak, Kornelia
10    (ii) TITLE OF INVENTION: ISOLATED p27 PROTEIN AND METHOD FOR ITS
11                               PRODUCTION AND USE
13    (iii) NUMBER OF SEQUENCES: 27
15    (iv) CORRESPONDENCE ADDRESS:
16          (A) ADDRESSEE: FOLEY, HOAG & ELIOT LLP
17          (B) STREET: One Post Office Square
18          (C) CITY: Boston
19          (D) STATE: MA
20          (E) COUNTRY: USA
21          (F) ZIP: 02109-2170
23    (v) COMPUTER READABLE FORM:
24          (A) MEDIUM TYPE: Floppy disk
25          (B) COMPUTER: IBM PC compatible
26          (C) OPERATING SYSTEM: PC-DOS/MS-DOS
27          (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
29    (vi) CURRENT APPLICATION DATA:
C--> 30          (A) APPLICATION NUMBER: US/09/865,018
C--> 31          (B) FILING DATE: 24-May-2001
33    (vii) PRIOR APPLICATION DATA:
34          (A) APPLICATION NUMBER: US/08/854,039
35          (B) FILING DATE: 09-MAY-1997
37    (viii) ATTORNEY/AGENT INFORMATION:
38          (A) NAME: Vincent, Matthew P.
39          (B) REGISTRATION NUMBER: 36,709
40          (C) REFERENCE/DOCKET NUMBER: MIV-079.04
42    (ix) TELECOMMUNICATION INFORMATION:
43          (A) TELEPHONE: 617-832-1000
44          (B) TELEFAX: 617-832-7000
47 (2) INFORMATION FOR SEQ ID NO: 1:
49     (i) SEQUENCE CHARACTERISTICS:
50          (A) LENGTH: 596 base pairs
51          (B) TYPE: nucleic acid
52          (C) STRANDEDNESS: single
53          (D) TOPOLOGY: linear
55     (ii) MOLECULE TYPE: cDNA
58     (ix) FEATURE:
59          (A) NAME/KEY: CDS
60          (B) LOCATION: 1..594
63     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
65 ATG TCA AAC GTG CGA GTG TCT AAC GGG AGC CCT AGC CTG GAG CGG ATG 48
66 Met Ser Asn Val Arg Val Ser Asn Gly Ser Pro Ser Leu Glu Arg Met

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/865,018

DATE: 11/29/2001

TIME: 15:58:29

Input Set : A:\09865018.raw

Output Set: N:\CRF3\11292001\I865018.raw

```

67   1           5           10           15
69 GAC GCC AGG CAG GCG GAG CAG CCC AAG CCC TCG GCC TGC AGG AAC CTC 96
70 Asp Ala Arg Gln Ala Glu Gln Pro Lys Pro Ser Ala Cys Arg Asn Leu
71           20           25           30
73 TTC GGC CCG GTG GAC CAC GAA GAG TTA ACC CGG GAC TTG GAG AAG CAC 144
74 Phe Gly Pro Val Asp His Glu Glu Leu Thr Arg Asp Leu Glu Lys His
75           35           40           45
77 TGC AGA GAC ATG GAA GAG GCG AGC CAG CGC AAG TGG AAT TTC GAT TTT 192
78 Cys Arg Asp Met Glu Glu Ala Ser Gln Arg Lys Trp Asn Phe Asp Phe
79           50           55           60
81 CAG AAT CAC AAA CCC CTA GAG GGC AAG TAC GAG TGG CAA GAG GTG GAG 240
82 Gln Asn His Lys Pro Leu Glu Gly Lys Tyr Glu Trp Gln Glu Val Glu
83 65           70           75           80
85 AAG GGC AGC TTG CCC GAG TTC TAC TAC AGA CCC CCG CGG CCC CCC AAA 288
86 Lys Gly Ser Leu Pro Glu Phe Tyr Tyr Arg Pro Pro Arg Pro Pro Lys
87           85           90           95
89 GGT GGC TGC AAG GTG CCG GCG CAG GAG AGC CAG GAT GTC AGC GGG AGC 336
90 Gly Gly Cys Lys Val Pro Ala Gln Glu Ser Gln Asp Val Ser Gly Ser
91           100           105           110
93 CGC CCG GCG GCG CCT TTA ATT GGG GCT CCG GCT AAC TCT GAG GAC ACG 384
94 Arg Pro Ala Ala Pro Leu Ile Gly Ala Pro Ala Asn Ser Glu Asp Thr
95           115           120           125
98 CAT TTG GTG GAC CCA AAG ACT GAT CCG TCG GAC AGC CAG ACG GGG TTA 432
99 His Leu Val Asp Pro Lys Thr Asp Pro Ser Asp Ser Gln Thr Gly Leu
100           130           135           140
102 GCG GAG CAA TGC GCA GGA ATA AGG AAG CGA CCT GCA ACC GAC GAT TCT 480
103 Ala Glu Gln Cys Ala Gly Ile Arg Lys Arg Pro Ala Thr Asp Asp Ser
104 145           150           155           160
106 TCT ACT CAA AAC AAA AGA GCC AAC AGA ACA GAA GAA AAT GTT TCA GAC 528
107 Ser Thr Gln Asn Lys Arg Ala Asn Arg Thr Glu Glu Asn Val Ser Asp
108           165           170           175
110 GGT TCC CCA AAT GCC GGT TCT GTG GAG CAG ACG CCC AAG AAG CCT GGC 576
111 Gly Ser Pro Asn Ala Gly Ser Val Glu Gln Thr Pro Lys Lys Pro Gly
112           180           185           190
114 CTC AGA AGA CGT CAA ACG TA 596
115 Leu Arg Arg Arg Gln Thr
116           195
119 (2) INFORMATION FOR SEQ ID NO: 2:
121 (i) SEQUENCE CHARACTERISTICS:
122 (A) LENGTH: 198 amino acids
123 (B) TYPE: amino acid
124 (D) TOPOLOGY: linear
127 (ii) MOLECULE TYPE: protein
130 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
132 Met Ser Asn Val Arg Val Ser Asn Gly Ser Pro Ser Leu Glu Arg Met
133 1           5           10           15
135 Asp Ala Arg Gln Ala Glu Gln Pro Lys Pro Ser Ala Cys Arg Asn Leu
136           20           25           30
138 Phe Gly Pro Val Asp His Glu Glu Leu Thr Arg Asp Leu Glu Lys His

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/865,018

DATE: 11/29/2001

TIME: 15:58:29

Input Set : A:\09865018.raw

Output Set: N:\CRF3\11292001\I865018.raw

```

139          35          40          45
141 Cys Arg Asp Met Glu Glu Ala Ser Gln Arg Lys Trp Asn Phe Asp Phe
142          50          55          60
144 Gln Asn His Lys Pro Leu Glu Gly Lys Tyr Glu Trp Gln Glu Val Glu
145 65          70          75          80
147 Lys Gly Ser Leu Pro Glu Phe Tyr Tyr Arg Pro Pro Arg Pro Pro Lys
148          85          90          95
150 Gly Gly Cys Lys Val Pro Ala Gln Glu Ser Gln Asp Val Ser Gly Ser
151          100          105          110
153 Arg Pro Ala Ala Pro Leu Ile Gly Ala Pro Ala Asn Ser Glu Asp Thr
154          115          120          125
156 His Leu Val Asp Pro Lys Thr Asp Pro Ser Asp Ser Gln Thr Gly Leu
157          130          135          140
159 Ala Glu Gln Cys Ala Gly Ile Arg Lys Arg Pro Ala Thr Asp Asp Ser
160 145          150          155          160
162 Ser Thr Gln Asn Lys Arg Ala Asn Arg Thr Glu Glu Asn Val Ser Asp
163          165          170          175
165 Gly Ser Pro Asn Ala Gly Ser Val Glu Gln Thr Pro Lys Lys Pro Gly
166          180          185          190
168 Leu Arg Arg Arg Gln Thr
169          195
172 (2) INFORMATION FOR SEQ ID NO: 3:
174 (i) SEQUENCE CHARACTERISTICS:
175 (A) LENGTH: 593 base pairs
176 (B) TYPE: nucleic acid
177 (C) STRANDEDNESS: single
178 (D) TOPOLOGY: linear
181 (ii) MOLECULE TYPE: cDNA
184 (ix) FEATURE:
185 (A) NAME/KEY: CDS
186 (B) LOCATION: 1..591
189 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
191 ATG TCA AAC GTG AGA GTG TCT AAC GGG AGC CCG AGC CTG GAG CGG ATG 48
192 Met Ser Asn Val Arg Val Ser Asn Gly Ser Pro Ser Leu Glu Arg Met
193 1 5 10 15
195 GAC GCC AGA CAA GCG GAT CAC CCC AAG CCT TCC GCC TGC AGA AAT CTC 96
196 Asp Ala Arg Gln Ala Asp His Pro Lys Pro Ser Ala Cys Arg Asn Leu
197 20 25 30
199 TTC GGC CCG GTC AAT CAT GAA GAA CTA ACC CGG GAC TTG GAG AAG CAC 144
200 Phe Gly Pro Val Asn His Glu Glu Leu Thr Arg Asp Leu Glu Lys His
201 35 40 45
203 TGC CGG GAT ATG GAA GAA GCG AGT CAG CGC AAG TGG AAT TTC GAC TTT 192
204 Cys Arg Asp Met Glu Glu Ala Ser Gln Arg Lys Trp Asn Phe Asp Phe
205 50 55 60
207 CAG AAT CAT AAG CCC CTG GAG GGC AGA TAC GAA TGG CAG GAG GTG GAG 240
208 Gln Asn His Lys Pro Leu Glu Gly Arg Tyr Glu Trp Gln Glu Val Glu
209 65 70 75 80
211 AGG GGC AGC TTG CCC GAG TTC TAC TAC AGG CCC CCG CGC CCC CCC AAG 288
212 Arg Gly Ser Leu Pro Glu Phe Tyr Tyr Arg Pro Pro Arg Pro Pro Lys

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/865,018

DATE: 11/29/2001

TIME: 15:58:29

Input Set : A:\09865018.raw

Output Set: N:\CRF3\11292001\I865018.raw

```

213          85          90          95
215 AGC GCC TGC AAG GTG CTG GCG CAG GAG AGC CAG GAT GTC AGC GGG AGC 336
216 Ser Ala Cys Lys Val Leu Ala Gln Glu Ser Gln Asp Val Ser Gly Ser
217          100          105          110
219 CGC CAG GCG GTG CCT TTA ATT GGG TCT CAG GCA AAC TCT GAG GAC CGG 384
220 Arg Gln Ala Val Pro Leu Ile Gly Ser Gln Ala Asn Ser Glu Asp Arg
221          115          120          125
223 CAT TTG GTG GAC CAA ATG CCT GAC TCG TCA GAC ATT CAG GCT GGG TTA 432
224 His Leu Val Asp Gln Met Pro Asp Ser Ser Asp Ile Gln Ala Gly Leu
225          130          135          140
227 GCG GAG CAG TGT CCA GGG ATG AGG AAG CGA CCT GCT GCA GAA GAT TCT 480
228 Ala Glu Gln Cys Pro Gly Met Arg Lys Arg Pro Ala Ala Glu Asp Ser
229 145          150          155          160
231 TCT TCG CAA AAC AAA AGG GCC AAC AGA ACA GAA GAA AAT GTT TCA GAC 528
232 Ser Ser Gln Asn Lys Arg Ala Asn Arg Thr Glu Glu Asn Val Ser Asp
233          165          170          175
235 GGT TCC CCG AAC GCT GGC ACT GTG GAG CAG ACG CCC AAG AAG CCC GGC 576
236 Gly Ser Pro Asn Ala Gly Thr Val Glu Gln Thr Pro Lys Lys Pro Gly
237          180          185          190
239 CTT CGA CGC CAG ACG TA 593
240 Leu Arg Arg Gln Thr
241          195
244 (2) INFORMATION FOR SEQ ID NO: 4:
246 (i) SEQUENCE CHARACTERISTICS:
247 (A) LENGTH: 197 amino acids
248 (B) TYPE: amino acid
249 (D) TOPOLOGY: linear
252 (ii) MOLECULE TYPE: protein
254 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
256 Met Ser Asn Val Arg Val Ser Asn Gly Ser Pro Ser Leu Glu Arg Met
257 1 5 10 15
259 Asp Ala Arg Gln Ala Asp His Pro Lys Pro Ser Ala Cys Arg Asn Leu
260 20 25 30
262 Phe Gly Pro Val Asn His Glu Glu Leu Thr Arg Asp Leu Glu Lys His
263 35 40 45
265 Cys Arg Asp Met Glu Glu Ala Ser Gln Arg Lys Trp Asn Phe Asp Phe
266 50 55 60
268 Gln Asn His Lys Pro Leu Glu Gly Arg Tyr Glu Trp Gln Glu Val Glu
269 65 70 75 80
271 Arg Gly Ser Leu Pro Glu Phe Tyr Tyr Arg Pro Pro Arg Pro Pro Lys
272 85 90 95
274 Ser Ala Cys Lys Val Leu Ala Gln Glu Ser Gln Asp Val Ser Gly Ser
275 100 105 110
277 Arg Gln Ala Val Pro Leu Ile Gly Ser Gln Ala Asn Ser Glu Asp Arg
278 115 120 125
280 His Leu Val Asp Gln Met Pro Asp Ser Ser Asp Ile Gln Ala Gly Leu
281 130 135 140
283 Ala Glu Gln Cys Pro Gly Met Arg Lys Arg Pro Ala Ala Glu Asp Ser
284 145 150 155 160

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/865,018

DATE: 11/29/2001

TIME: 15:58:29

Input Set : A:\09865018.raw

Output Set: N:\CRF3\11292001\I865018.raw

```

286 Ser Ser Gln Asn Lys Arg Ala Asn Arg Thr Glu Glu Asn Val Ser Asp
287           165           170           175
289 Gly Ser Pro Asn Ala Gly Thr Val Glu Gln Thr Pro Lys Lys Pro Gly
290           180           185           190
292 Leu Arg Arg Gln Thr
293           195
295 (2) INFORMATION FOR SEQ ID NO: 5:
297   (i) SEQUENCE CHARACTERISTICS:
298       (A) LENGTH: 534 base pairs
299       (B) TYPE: nucleic acid
300       (C) STRANDEDNESS: single
301       (D) TOPOLOGY: linear
303   (ii) MOLECULE TYPE: cDNA
306   (ix) FEATURE:
307       (A) NAME/KEY: CDS
308       (B) LOCATION: 1..534
311   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
313 ATG TCA AAC GTG CGG GTG TCT AAC GGG AGC CCG AGC CTG GAG CGG ATG 48
314 Met Ser Asn Val Arg Val Ser Asn Gly Ser Pro Ser Leu Glu Arg Met
315 1           5           10           15
317 GAC GCC AGA CAG GCG GAG TAC CCC AAG CCC TCC GCC TGC AGA AAC CTC 96
318 Asp Ala Arg Gln Ala Glu Tyr Pro Lys Pro Ser Ala Cys Arg Asn Leu
319           20           25           30
321 TTC GGC CCG GTC AAC CAC GAA GAG CTG ACC CGG GAC TTG GAG AAG CAC 144
322 Phe Gly Pro Val Asn His Glu Glu Leu Thr Arg Asp Leu Glu Lys His
323           35           40           45
325 CGC AGA GAC ATG GAA GAG GCA AGC CAG CGC AAG TGG AAT TTT GAT TTC 192
326 Arg Arg Asp Met Glu Glu Ala Ser Gln Arg Lys Trp Asn Phe Asp Phe
327           50           55           60
329 CAG AAT CAC AAG CCC CTG GAG GGC AAA TAC GAG TGG CAG GAG GTG GAG 240
330 Gln Asn His Lys Pro Leu Glu Gly Lys Tyr Glu Trp Gln Glu Val Glu
331 65           70           75           80
333 AAG GGC AGC TTG CCG GAG TTC TAC TAC AGA CCC CCG CGG CCA CCC AAA 288
334 Lys Gly Ser Leu Pro Glu Phe Tyr Tyr Arg Pro Pro Arg Pro Pro Lys
335           85           90           95
337 GGC GCC TGC AAG GTG CCG GCG CAG GAG AGC CAG GAC GTC AGC GGG ACC 336
338 Gly Ala Cys Lys Val Pro Ala Gln Glu Ser Gln Asp Val Ser Gly Thr
339           100           105           110
341 CGG CAG GCC GTG CCT TTA ATG GGG TCT CAG GCA AAC TCA GAG GAC ACA 384
342 Arg Gln Ala Val Pro Leu Met Gly Ser Gln Ala Asn Ser Glu Asp Thr
343           115           120           125
345 CAC TTG GTA GAC CAA AAG ACT GAC ACG GCG GAC AAC CAG GCT GGC TTA 432
346 His Leu Val Asp Gln Lys Thr Asp Thr Ala Asp Asn Gln Ala Gly Leu
347           130           135           140
349 GCG GAG CAG TGC ACT GGG ATC AGG AAG CGA CCG GCC ACA GAC GAT TCC 480
350 Ala Glu Gln Cys Thr Gly Ile Arg Lys Arg Pro Ala Thr Asp Asp Ser
351 145           150           155           160
353 TCT CCT CAA AAC AAA AGA GCC AAC AGA ACA GAA GAA AAT GTC TCA GAC 528
354 Ser Pro Gln Asn Lys Arg Ala Asn Arg Thr Glu Glu Asn Val Ser Asp

```

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/09/865,018

DATE: 11/29/2001

TIME: 15:58:30

Input Set : A:\09865018.raw

Output Set: N:\CRF3\11292001\I865018.raw

L:30 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:31 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]